

# PHASE DETECTION DEVICE AND METHOD AND SPEECH ENCODING DEVICE AND METHOD

**Patent number:** JP11219198  
**Publication date:** 1999-08-10  
**Inventor:** INOUE AKIRA; NISHIGUCHI MASAYUKI  
**Applicant:** SONY CORP  
**Classification:**  
 - International: G10L9/14; G10L9/16; H03M7/30  
 - european:  
**Application number:** JP19980019963 19980130  
**Priority number(s):** JP19980019963 19980130

Also published as:



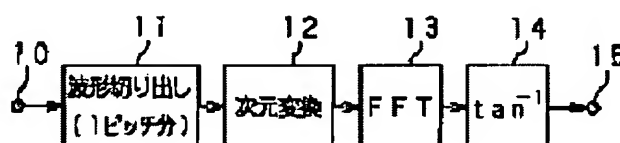
US6115685 (A1)

Report a data error here

## Abstract of JP11219198

**PROBLEM TO BE SOLVED:** To detect the phase information of an input signal, for example, at the time of synthesizing and encoding sine waves without any need of interpolation after FFT.

**SOLUTION:** The waveform of an input signal on the basis of a speech signal from an input terminal 10 is clipped out at a waveform clipping part 11 by one-pitch period on a time axis, and dimensionally converted into  $2^k$  samples ( $k$ : integer) through dimensional conversion 12. Waveform data so converted dimensionally is subjected to an FFT process at an FFT (Fast Fourier Transform) part 13. Then,  $\tan^{-1}$  is calculated at a  $\tan^{-1}$  part 14 using the real part and the imaginary part of the FFT processed data, thereby directly finding the phase of an input signal for every harmonics.



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 ( J P )

(12) 公開特許公報 ( A )

(11) 特許出願公開番号

特開平 1 1 - 2 1 9 1 9 8

(43) 公開日 平成 1 1 年 ( 1 9 9 9 ) 8 月 1 0 日

(51) Int. Cl.	識別記号	庁内整理番号	F I	技術表示箇所
G10L 9/14			G10L 9/14	G
9/16			9/16	
// H03M 7/30			H03M 7/30	A

審査請求 未請求 請求項の数 1 8 O L (全 1 2 頁)

(21) 出願番号 特願平 1 0 - 1 9 9 6 3

(22) 出願日 平成 1 0 年 ( 1 9 9 8 ) 1 月 3 0 日

(71) 出願人 0 0 0 0 0 2 1 8 5

ソニー株式会社

東京都品川区北品川 6 丁目 7 番 3 5 号

(72) 発明者 井上 晃

東京都品川区北品川 6 丁目 7 番 3 5 号 ソ

ニー株式会社内

(72) 発明者 西口 正之

東京都品川区北品川 6 丁目 7 番 3 5 号 ソ

ニー株式会社内

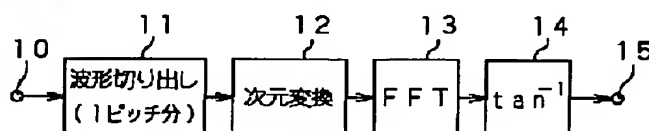
(74) 代理人 弁理士 小池 晃 (外 2 名)

(54) 【発明の名称】 位相検出装置及び方法、並びに音声符号化装置及び方法

(57) 【要約】

【課題】 サイン波合成符号化の際等の入力信号の位相情報を、FFT後の補間処理なしに検出する。

【解決手段】 入力端子 1 0 からの音声信号に基づく入力信号の波形を波形切り出し部 1 1 で時間軸上で 1 ピッチ周期分だけ切り出し、次元変換 1 2 により  $2^k$  サンプル (k は整数) に次元変換する。この次元変換された波形データに対して FFT (高速フーリエ変換) 部 1 3 で FFT 処理し、FFT 処理されたデータの実部と虚部とを用いて  $\tan^{-1}$  部 1 4 で  $\tan^{-1}$  の計算を行うことにより、直接的に入力信号の各高調波毎の位相を求める。



## 【特許請求の範囲】

【請求項 1】 音声信号に基づく入力信号波形を時間軸上で 1 ピッチ周期分だけ切り出す波形切り出し手段と、切り出された 1 ピッチ周期分の波形データのサンプル数を  $2^k$  サンプル ( $k$  は整数) に次元変換する次元変換手段と、

この次元変換された  $2^k$  サンプルのデータに対して直交変換を施す直交変換手段と、

この直交変換手段からのデータの実部と虚部とに基づいて上記入力信号の各高調波成分の位相情報を検出する位相検出手段とを有することを特徴とする位相検出装置。

【請求項 2】 上記入力信号波形は音声信号波形であることを特徴とする請求項 1 記載の位相検出装置。

【請求項 3】 上記入力信号波形は音声信号の短期予測残差の信号波形であることを特徴とする請求項 1 記載の位相検出装置。

【請求項 4】 上記次元変換手段は、上記波形切り出し手段からの切り出し波形データに対してオーバーサンプリング、線形補間を施して  $2^k$  サンプルのデータに次元変換することを特徴とする請求項 1 記載の位相検出装置。

【請求項 5】 上記直交変換手段は、高速フーリエ変換回路であり、上記次元変換された  $2^k$  サンプルのデータに対して  $2^k$  ポイント高速フーリエ変換処理を施すことを特徴とする請求項 1 記載の位相検出装置。

【請求項 6】 上記位相検出手段は、上記直交変換手段からのデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により各高調波毎の位相を求めることを特徴とする請求項 1 記載の位相検出装置。

【請求項 7】 音声信号に基づく入力信号波形を時間軸上で 1 ピッチ周期分だけ切り出す波形切り出し工程と、切り出された 1 ピッチ周期分の波形データのサンプル数を  $2^k$  サンプル ( $k$  は整数) に次元変換する次元変換工程と、

この次元変換された  $2^k$  サンプルのデータに対して直交変換を施す直交変換工程と、

この直交変換工程により得られたデータの実部と虚部とに基づいて上記入力信号の各高調波成分の位相情報を検出する位相検出工程とを有することを特徴とする位相検出方法。

【請求項 8】 上記次元変換工程は、上記波形切り出し工程により得られた切り出し波形データに対してオーバーサンプリング、線形補間を施して  $2^k$  サンプルのデータに次元変換することを特徴とする請求項 7 記載の位相検出方法。

【請求項 9】 上記位相検出工程では、上記直交変換工程により得られたデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により各高調波毎の位相を求めることを特徴とする請求項 7 記載の位相検出方法。

【請求項 10】 音声信号に基づく入力信号を時間軸上でブロック単位で区分し、区分された各ブロック毎にピ

ッチを求めると共に、各ブロック単位でサイン波分析合成符号化を施す音声符号化装置において、

上記入力信号の波形を時間軸上で上記ピッチの 1 ピッチ周期分だけ切り出す波形切り出し手段と、

切り出された 1 ピッチ周期分の波形データのサンプル数を  $2^k$  サンプル ( $k$  は整数、 $2^k$  は上記 1 ピッチ周期のサンプル数以上) に次元変換する次元変換手段と、

この次元変換された  $2^k$  サンプルのデータに対して直交変換を施す直交変換手段と、

この直交変換手段からのデータの実部と虚部とに基づいて上記入力信号の上記サイン波合成のための各高調波成分の位相情報を検出する位相検出手段とを有することを特徴とする音声符号化装置。

【請求項 11】 上記入力信号は音声信号であることを特徴とする請求項 10 記載の音声符号化装置。

【請求項 12】 上記入力信号は音声信号の短期予測残差信号であることを特徴とする請求項 10 記載の音声符号化装置。

【請求項 13】 上記次元変換手段は、上記波形切り出し手段からの切り出し波形データに対してオーバーサンプリング、線形補間を施して  $2^k$  サンプルのデータに次元変換することを特徴とする請求項 10 記載の音声符号化装置。

【請求項 14】 上記直交変換手段は、高速フーリエ変換回路であり、上記次元変換された  $2^k$  サンプルのデータに対して  $2^k$  ポイント高速フーリエ変換処理を施すことを特徴とする請求項 10 記載の音声符号化装置。

【請求項 15】 上記位相検出手段は、上記直交変換手段からのデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により各高調波毎の位相を求めることを特徴とする請求項 10 記載の音声符号化装置。

【請求項 16】 音声信号に基づく入力信号を時間軸上でブロック単位で区分し、区分された各ブロック毎にピッチを求めると共に、各ブロック単位でサイン波分析合成符号化を施す音声符号化方法において、上記入力信号の波形を時間軸上で上記ピッチの 1 ピッチ周期分だけ切り出す波形切り出し工程と、

切り出された 1 ピッチ周期分の波形データのサンプル数を  $2^k$  サンプル ( $k$  は整数) に次元変換する次元変換工程と、

この次元変換された  $2^k$  サンプルのデータに対して直交変換を施す直交変換工程と、

この直交変換工程により得られたデータの実部と虚部とに基づいて上記入力信号の各高調波成分の位相情報を検出する位相検出工程とを有することを特徴とする音声符号化方法。

【請求項 17】 上記次元変換工程は、上記波形切り出し工程により得られた切り出し波形データに対してオーバーサンプリング、線形補間を施して  $2^k$  サンプルのデータに次元変換することを特徴とする請求項 16 記載の音

声符号化方法。

【請求項 1 8】 上記位相検出工程では、上記直交変換工程により得られたデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により各高調波毎の位相を求めることを特徴とする請求項 1 6 記載の音声符号化方法。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】 本発明は、サイン波合成符号化等における各高調波（ハーモニクス）成分の位相を検出するための位相検出装置及び方法、並びに音声符号化装置及び方法に関する。

【 0 0 0 2 】

【従来の技術】 オーディオ信号（音声信号や音響信号を含む）の時間領域や周波数領域における統計的性質と人間の聴感上の特性を利用して信号圧縮を行うような符号化方法が種々知られている。この符号化方法としては、大別して時間領域での符号化、周波数領域での符号化、分析合成符号化等が挙げられる。

【 0 0 0 3 】 音声信号等の高能率符号化の例としては、ハーモニック（Harmonic）符号化、M B E（Multiband Excitation: マルチバンド励起）符号化等のサイン波分析合成符号化（Sinusoidal Coding）や、S B C（Sub-band Coding: 帯域分割符号化）、L P C（Linear Predictive Coding: 線形予測符号化）、あるいは D C T（離散コサイン変換）、M D C T（モデファイド D C T）、F F T（高速フーリエ変換）等が知られている。

【 0 0 0 4 】

【発明が解決しようとする課題】 ところで、入力音声信号に対して上記 M B E 符号化、ハーモニック符号化や、S T C（Sinusoidal Transform Coding）等のサイン波合成符号化（Sinusoidal Coding）を用いるような、又は、入力音声信号の L P C（線形予測符号化）残差に対してこれらのサイン波合成符号化を用いるような音声高能率符号化においては、分析合成の要素となる各サイン波（ハーモニクス、高調波）の振幅、あるいはスペクトルエンベロープに関する情報を伝送しているが、位相については伝送しておらず、合成時に適宜に位相を算出しているのが実情である。

【 0 0 0 5 】 そのため、復号されて再生される音声波形は、元の入力音声信号の波形と異なることになる、という問題がある。すなわち、元の波形の波形再生を実現するためには、各ハーモニクス（高調波）成分の位相情報をフレーム毎に検出して伝送することが必要とされる。

【 0 0 0 6 】 本発明は、このような実情に鑑みてなされたものであり、元の波形の波形再現性を実現するための位相検出装置及び方法、並びにこの位相検出の技術を用いた音声符号化装置及び方法の提供を目的とする。

【 0 0 0 7 】

【課題を解決するための手段】 本発明に係る位相検出装

置及び方法は、上述した課題を解決するために、音声信号に基づく入力信号波形を時間軸上で 1 ピッチ周期分だけ切り出し、切り出された 1 ピッチ周期分のサンプルに対して次元変換を施して  $2^k$  サンプル（ $k$  は整数、 $2^k$  は上記 1 ピッチ周期のサンプル数以上）のデータにし、この次元変換されたデータに対して  $2^k$  ポイント F F T 等の直交変換を施し、直交変換されたデータの実部と虚部とに基づいて上記入力信号の各高調波成分の位相情報を検出することを中心としている。

【 0 0 0 8 】 また、本発明は、上記特徴を有する位相検出を、サイン波合成符号化等の音声符号化に適用することを特徴としている。

【 0 0 0 9 】 ここで、上記入力信号波形としては、音声信号波形そのもの、あるいは音声信号の短期予測残差の信号波形を用いることができる。

【 0 0 1 0 】 また、上記次元変換としては、上記切り出された波形データに対してオーバーサンプリング、線形補間を施して  $2^k$  サンプルのデータに変換することが挙げられる。

【 0 0 1 1 】 さらに、上記位相検出は、上記直交変換により得られたデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により各高調波毎の位相を求めることが挙げられる。

【 0 0 1 2 】

【発明の実施の形態】 本発明に係る位相検出装置及び方法は、例えばマルチバンド励起（Multiband Excitation: M B E）符号化、サイン波変換符号化（Sinusoidal Transform Coding: S T C）、ハーモニック符号化（Harmonic coding）等のサイン波合成符号化方式に適用されるものであり、又は L P C（Linear Predictive Coding）残差に上記サイン波合成符号化を用いた符号化方式に適用されるものである。

【 0 0 1 3 】 ここで、本発明の実施の形態の説明に先立ち、本発明に係る位相検出装置あるいは方法が適用される装置としてのサイン波分析合成符号化を行うような音声符号化装置について説明する。

【 0 0 1 4 】 図 1 は、上述した位相検出装置あるいは方法が適用される音声符号化装置の具体例の概略構成を示している。

【 0 0 1 5 】 図 1 の音声符号化装置は、入力信号に対して、サイン波分析（sinusoidal analysis）符号化、例えばハーモニックコーディング（harmonic coding）を行う第 1 の符号化部 1 1 0 と、入力信号に対して、例えば合成による分析法を用いて最適ベクトルのクローズドループサーチによるベクトル量子化を用いた符号励起線形予測（C E L P）符号化を施す第 2 の符号化部 1 2 0 とを有し、入力信号の有声音（V: Voiced）の部分の符号化に第 1 の符号化部 1 1 0 を用い、入力信号の無声音（U V: Unvoiced）の部分の符号化には第 2 の符号化部 1 2 0 を用いるようにしている。本発明に係る

位相検出の実施の形態は、第 1 の符号化部 1 1 0 に対して適用されている。なお、図 1 の例では、入力音声信号の短期予測残差例えば L P C (線形予測符号化) 残差を求めた後に第 1 の符号化部 1 1 0 に送られるようにしている。

【 0 0 1 6 】図 1 において、入力端子 1 0 1 に供給された音声信号は、L P C 逆フィルタ 1 3 1 及び L P C 分析部 1 3 2 に送られ、また、第 1 の符号化部 1 1 0 のオープンループピッチサーチ部 1 1 1 にも送られる。L P C 分析部 1 3 2 は、入力信号波形の 2 5 6 サンプル程度の長さ (分析長) を 1 ブロックとしてハミング窓をかけて、自己相関法により線形予測係数、いわゆる  $\alpha$  パラメータを求める。データ出力の単位となるフレーミングの間隔は、1 6 0 サンプル程度とする。ここで、入力音声信号のサンプリング周波数  $f_s$  が例えば 8 k H z のとき、1 フレーム間隔は 1 6 0 サンプルで 2 0 m s e c となる。

【 0 0 1 7 】L P C 分析部 1 3 2 からの  $\alpha$  パラメータは、例えば  $\alpha \rightarrow$  L S P 変換により線スペクトル対 (L S P) パラメータに変換される。これは、直接型のフィルタ係数として求めた  $\alpha$  パラメータを、例えば 1 0 個、すなわち 5 対の L S P パラメータに変換する。変換は例えばニュートン-ラプソン法等を用いて行う。この L S P パラメータに変換するのは、 $\alpha$  パラメータよりも補間特性に優れているからである。この L S P パラメータは、L S P 量子化器 1 3 3 によりマトリクスあるいはベクトル量子化される。このとき、フレーム間差分をとってからベクトル量子化してもよく、複数フレーム分をまとめてマトリクス量子化してもよい。ここでは、2 0 m s e c を 1 フレームとし、2 0 m s e c 毎に算出される L S P パラメータを 2 フレーム分まとめて、マトリクス量子化及びベクトル量子化している。

【 0 0 1 8 】この L S P 量子化器 1 3 3 からの量子化出力、すなわち L S P 量子化のインデックスは、端子 1 0 2 を介して取り出され、また量子化済みの L S P ベクトルは、例えば L S P 補間や L S P  $\rightarrow \alpha$  変換を介して L P C の  $\alpha$  パラメータとされて、L P C 逆フィルタ 1 3 1 や、後述する第 2 の符号化部 1 2 0 の聴覚重み付きの L P C 合成フィルタ 1 2 2 及び聴覚重み付けフィルタ 1 2 5 に送られる。

【 0 0 1 9 】また、L P C 分析部 1 3 2 からの  $\alpha$  パラメータは、聴覚重み付けフィルタ算出部 1 3 4 に送られて聴覚重み付けのためのデータが求められ、この重み付けデータが後述する聴覚重み付きのベクトル量子化器 1 1 6 と、第 2 の符号化部 1 2 0 の聴覚重み付きの L P C 合成フィルタ 1 2 2 及び聴覚重み付けフィルタ 1 2 5 とに送られる。

【 0 0 2 0 】L P C 逆フィルタ 1 3 1 では、上記  $\alpha$  パラメータを用いて、入力音声信号の線形予測残差 (L P C 残差) を取り出すような逆フィルタリング処理を行っている。この L P C 逆フィルタ 1 3 1 からの出力は、サイ

ン波分析符号化、具体的には例えばハーモニック符号化を行う第 1 の符号化部 1 1 0 の、D F T (離散フーリエ変換) 回路等の直交変換部 1 1 2 及び位相検出部 1 4 0 に送られる。

【 0 0 2 1 】また、符号化部 1 1 0 のオープンループピッチサーチ部 1 1 1 には、上記入力端子 1 0 1 からの入力音声信号が供給されている。オープンループピッチサーチ部 1 1 1 では、入力信号の L P C 残差をとってオープンループによる比較的ラフなピッチのサーチが行われ、抽出された粗ピッチデータは高精度ピッチサーチ部 1 1 3 に送られて、後述するようなクロズドループによる高精度のピッチサーチ (ピッチのファインサーチ) が行われる。また、オープンループピッチサーチ部 1 1 1 からは、上記粗ピッチデータと共に L P C 残差の自己相関の最大値をパワーで正規化した正規化自己相関最大値  $r(p)$  が取り出され、 $V/U V$  (有声音/無声音) 判定部 1 1 4 に送られている。

【 0 0 2 2 】直交変換部 1 1 2 では例えば D F T (離散フーリエ変換) 等の直交変換処理が施されて、時間軸上の L P C 残差が周波数軸上のスペクトル振幅データに変換される。この直交変換部 1 1 2 からの出力は、高精度ピッチサーチ部 1 1 3 及びスペクトル振幅あるいはエンベロープを評価するためのスペクトルエンベロープ評価部 1 1 5 に送られる。

【 0 0 2 3 】高精度 (ファイン) ピッチサーチ部 1 1 3 には、オープンループピッチサーチ部 1 1 1 で抽出された比較的ラフな粗ピッチデータと、直交変換部 1 1 2 により例えば D F T された周波数軸上のデータとが供給されている。この高精度ピッチサーチ部 1 1 3 では、上記粗ピッチデータ値を中心に、0.2 ~ 0.5 きざみで  $\pm$  数サンプルずつ振って、最適な小数点付き (フローティング) のファインピッチデータの値へ追い込む。このときのファインサーチの手法として、いわゆる合成による分析 (Analysis by Synthesis) 法を用い、合成されたパワースペクトルが原音のパワースペクトルに最も近くなるようにピッチを選んでいる。このようなクロズドループによる高精度のピッチサーチ部 1 1 4 6 からのピッチデータについては、スペクトルエンベロープ評価部 1 1 5、位相検出部 1 4 1、及び切換部 1 0 7 に送っている。

【 0 0 2 4 】スペクトルエンベロープ評価部 1 1 5 では、L P C 残差の直交変換出力としてのスペクトル振幅及びピッチに基づいて各ハーモニクスの大きさ及びその集合であるスペクトルエンベロープが評価され、高精度ピッチサーチ部 1 1 3、 $V/U V$  (有声音/無声音) 判定部 1 1 4 及びスペクトルエンベロープ量子化部 1 1 6 に送られる。スペクトルエンベロープ量子化部 1 1 6 としては、聴覚重み付きのベクトル量子化器が用いられる。

【 0 0 2 5 】 $V/U V$  (有声音/無声音) 判定部 1 1 4

は、直交変換部 1 1 2 からの出力と、高精度ピッチサーチ部 1 1 3 からの最適ピッチと、スペクトルエンベロープ評価部 1 1 5 からのスペクトル振幅データと、オープンループピッチサーチ部 1 1 1 からの正規化自己相関最大値  $r(p)$  とに基づいて、当該フレームの V/U V 判定が行われる。さらに、M B E の場合の各バンド毎の V/U V 判定結果の境界位置も当該フレームの V/U V 判定の一条件としてもよい。この V/U V 判定部 1 1 5 からの判定出力は、出力端子 1 0 5 を介して取り出される。

【 0 0 2 6 】ところで、スペクトル評価部 1 1 5 の出力部あるいはスペクトルエンベロープ量子化部 1 1 6 の入力部には、データ数変換（一種のサンプリングレート変換）部が設けられている。このデータ数変換部は、上記ピッチに応じて周波数軸上での分割帯域数が異なり、データ数が異なることを考慮して、エンベロープの振幅データ  $|A_i|$  を一定の個数にするためのものである。すなわち、例えば有効帯域を 3 4 0 0 k H z までとすると、この有効帯域が上記ピッチに応じて、8 バンド～6 3 バンドに分割されることになり、これらの各バンド毎に得られる上記振幅データ  $|A_i|$  の個数も 8 ～6 3 と変化することになる。このため上記データ数変換部で、この可変個数の振幅データを一定個数、例えば 4 4 個、のデータに変換している。

【 0 0 2 7 】このスペクトルエンベロープ評価部 1 1 5 の出力部あるいはスペクトルエンベロープ量子化部 1 1 6 の入力部に設けられたデータ数変換部からの上記一定個数（例えば 4 4 個）の振幅データあるいはエンベロープデータが、スペクトルエンベロープ量子化部 1 1 6 により、所定個数、例えば 4 4 個のデータ毎にまとめられてベクトルとされ、重み付きベクトル量子化が施される。この重みは、聴覚重み付けフィルタ算出回路 1 3 4 からの出力により与えられる。スペクトルエンベロープ量子化部 1 1 6 からの上記エンベロープのインデックスは、切換部 1 0 7 に送られる。

【 0 0 2 8 】位相検出部 1 4 1 では、後述するようにサイン波分析合成符号化の各ハーモニクス（高調波）毎の位相や位相の固定遅延成分等の位相情報を検出し、この位相情報を位相量子化部 1 4 2 に送って量子化し、量子化された位相データを切換部 1 0 7 に送っている。

【 0 0 2 9 】切換部 1 0 7 は、V/U V 判定部 1 1 5 からの V/U V 判定出力に応じて、第 1 の符号化部 1 1 0 のピッチ、スペクトルエンベロープのベクトル量子化インデックス、位相の各データと、第 2 の符号化部 1 2 0 からの後述するシェイプ、ゲインの各データとを切り換えて、端子 1 0 3 より出力する。

【 0 0 3 0 】図 1 の第 2 の符号化部 1 2 0 は、この例では C E L P（符号励起線形予測）符号化構成を有しており、雑音符号帳 1 2 1 からの出力を、重み付きの合成フィルタ 1 2 2 により合成処理し、得られた重み付き音声を減算器 1 2 3 に送り、入力端子 1 0 1 に供給された音

声信号を聴覚重み付けフィルタ 1 2 5 を介して得られた音声との誤差を取り出し、この誤差を距離計算回路 1 2 4 に送って距離計算を行い、誤差が最小となるようなベクトルを雑音符号帳 1 2 1 でサーチするような、合成による分析（Analysis by Synthesis）法を用いたクロズドループサーチを用いた時間軸波形のベクトル量子化を行っている。この C E L P 符号化は、上述したように無声音部分の符号化に用いられており、雑音符号帳 1 2 1 からの U V データとしてのコードブックインデックスは、上記 V/U V 判定部 1 1 5 からの V/U V 判定結果が無声音（U V）のとき切り換えられる切換部 1 0 7 を介して、出力端子 1 0 7 より取り出される。

【 0 0 3 1 】次に、本発明に係る好ましい実施の形態について、以下に説明する。この本発明に係る位相検出装置及び方法の実施の形態は、上記図 1 に示した音声信号符号化装置の位相検出部 1 4 1 に用いられるものであるが、これに限定されないことは勿論である。

【 0 0 3 2 】先ず、図 2 は、本発明に係る好ましい実施の形態となる位相検出装置の概略構成を示すブロック図、図 3 は、本発明に係る好ましい実施の形態となる位相検出方法を説明するためのフローチャートである。

【 0 0 3 3 】図 2 の入力端子 1 0 に供給される入力信号としては、デジタル化した音声信号そのもの、あるいは上述した図 1 の例の L P C 逆フィルタ 1 3 1 からの信号のようなデジタル音声信号の短期予測残差信号（L P C 残差信号）が用いられる。この入力信号に対して、波形切り出し部 1 1 により、図 3 のステップ S 1 1 に示すように、1 ピッチ周期分の波形信号を切り出している。これは、図 4 に示すように、入力信号（音声信号あるいは L P C 残差信号） $s(i)$  の分析ブロック中の分析点（時刻） $n$  から 1 ピッチ周期に相当するサンプル数（ピッチラグ） $pch$  を切り出す処理である。この図 4 の例では、分析ブロック長を 2 5 6 サンプルとしているが、これに限定されない。また、図 4 の横軸は分析ブロック中の位置あるいは時刻をサンプル数で表しており、上記分析点の位置あるいは時刻  $n$  は、分析開始から  $n$  サンプル目であることを示している。

【 0 0 3 4 】ここで、上記時刻  $n$ （サンプル）を中心とする分析ブロックのピッチラグが  $pch$ （サンプル）であるので、時刻  $n$  における基本周波数（角周波数） $\omega_0$  は、

$$\omega_0 = 2\pi / pch \quad (1)$$

となる。周波数軸上の  $\omega = 0 \sim \pi$  の範囲にハーモニクス（高調波）が  $\omega_0$  間隔に  $M$  本並んでいる。この  $M$  は、

$$M = pch / 2 \quad (2)$$

となる。

【 0 0 3 5 】この切り出された 1 ピッチ分の波形信号に対して、次元変換処理部 1 2 により、図 3 のステップ S 1 2 の次元変換処理が施される。これは、図 5 に示すように、上記 1 ピッチラグ分の  $pch$  サンプルの信号波形

を、 $2^7$  サンプル、この実施の形態では、 $2^7 = 128$  サンプルとなるように次元変換して、信号列  $re(i)$  (ただし、 $0 \leq i < 2^7$ ) を得る。この次元変換の具体例については後述する。

【0036】上記信号列  $re(i)$  を実数部とし、虚数信号列  $im(i)$  として、

$$im(i) = 0 \quad (0 \leq i < 2^7)$$

を用い、FFT処理部13により、図3のステップS13に示すように、これらの実数信号列  $re(i)$  及び虚数信号列  $im(i)$  に対して  $2^7$  ポイント、例えば128ポイントFFT(高速フーリエ変換)を実行する。

【0037】このFFTの実行結果に対して、 $\tan^{-1}$ 処理部14により、図3のステップS14に示すように  $\tan^{-1}$  (逆正接) を計算して位相を求める。これは、FFTの実行結果の実数部を  $Re(i)$ 、虚数部を  $Im(i)$  とするとき、 $0 \leq i < 2^7$  が周波数軸上の  $0 \sim \pi$  (rad) に相当し、これらの内の  $1 \leq i \leq M$  の成分が、上記M本の各ハーモニクス(高調波)の成分に相当する。すなわち、第m番目のハーモニクスの位相  $\phi_m = \phi(m \times \omega_s)$

(ただし、 $1 \leq m \leq M$ ) は、次の(3)式により表されることになる。

【0038】

【数1】

$$\phi_m = \tan^{-1} \left( \frac{Im(m)}{Re(m)} \right) \quad (1 \leq m \leq M) \quad (3)$$

【0039】求められた位相データは、出力端子15を介して取り出され、その具体例を図6に示す。この図6の×印が各ハーモニクスの位相を示しているが、上記

(3)式からも明らかなように、FFTの実行結果の各サンプルがそのまま各ハーモニクスの位相となってい

$$coef[i] = \frac{\sin \pi(i-32)/8}{\pi(i-32)} \left( 0.5 - 0.5 \cos 2\pi \frac{i-32}{64} \right) \quad 0 \leq i \leq 64 \quad (4)$$

【0044】図9及び図10は、この次元変換の処理の具体例を説明するためのものであり、図9は処理のためのメモリ内に蓄えられるデータの一例を示す図、図10はこの次元変換の処理手順を示すフローチャートである。

【0045】図9は、次元変換処理に先立って、メモリに蓄えられるデータを示しており、上記ピッチラグpchサンプルの前後に4サンプルずつ付加した次元数のバッファメモリに蓄えられる各変数buf(i)には、先頭4サンプルの領域に0が書き込まれ、次のpchサンプルの領域に上記入力信号から切り出された1ピッチ分の波形データ(次元変換前の波形データ)src(0)~src(pch-1)が書き込まれ、これに続く4サンプルの領域に上記1ピッチ分の波形データの最後のデータsrc(pch-1)が書き込まれる。また、次元変換処理後の  $2^7$  サンプル例えば128サンプルのデータは、128次元の変数dst(i)にdst(0)

る。このように、各ハーモニクスの位相がFFT実行結果に直接的に表れ、各ハーモニクスの周波数での位相を求めるための補間処理等が不要であることから、FFTのポイント数を少なくできるという利点がある。

【0040】例えば入力音声信号のサンプリング周波数を8kHzとすると、女声の高い方から男声の低い方まで、平均的なピッチラグpchは20~147(サンプル)程度である。従って、最低ピッチ周波数は、 $8000/147 \approx 54$  (Hz)であり、有効帯域を3.4kHzとすれば、最大62本程度のハーモニクスが並ぶことになる。従って、 $0 \sim \pi$ までの間に62ポイント以上の位相が求められればよいので、

$$62 \leq 2^7$$

を満足する整数値kの最小値が7であることより、上述したように128ポイントFFTを実行するだけで十分である。

【0041】次に、上記次元変換の具体例について説明する。この次元変換は、例えばオーバーサンプリングと線形補間により実現することができる。

【0042】例えば、入力音声信号のサンプリング周波数が8kHzのとき想定されるピッチラグpchは20~147(サンプル)程度であるので、1ピッチ波形を、 $2^7 = 128$  サンプルとなるように次元変換する。この場合、例えば図7のようなフィルタ係数を持ち、図8のような周波数特性を有するオーバーサンプリングフィルタを用いて8倍オーバーサンプリングし、必要な128個の点だけを線形補間で求める。上記オーバーサンプリングフィルタの図7に示すフィルタ係数coef(i)は、次の(4)式で表される。

【0043】

【数2】

~dst(127)として書き込まれる。

【0046】このようなメモリを用いて次元変換処理を行う場合、例えば図10のような8倍オーバーサンプリング処理フィルタ処理と補間処理とを組み合わせ、128サンプルのデータに変換することができる。この図10では、ステップS31にて変数ii、iを0とする初期化を行っており、変数iiを出力サンプルのカウントに、変数iは入力サンプルのカウントにそれぞれ用いている。ステップS32~S34では、iがpchを超えるまで、すなわち各サンプル毎の処理をステップS35でiをインクリメントしながら入力されたpchサンプルの処理を行うことを示しており、ステップS34~ステップS43は、入力pchサンプルにフィルタ係数coefを乗算して128サンプルの位置でのサンプル値を補間処理(ステップS41及びS42)により求める処理を示している。ステップS34、S36の固定値Rは8で、変

数  $p$  が 0 から 7 までステップ S 3 8 でインクリメントされながら処理を繰り返すような制御が行われる。ステップ S 3 6、S 3 7、S 4 3 における  $ip\_ratio$  や、ステップ S 3 9、S 4 1、S 4 3 における  $rel0, rel1$  は、補間のための入力サンプルに対する出力サンプルの関係や入力サンプル値の比例配分の数値を示すものである。この図 1 0 に示す次元変換の処理は、通常の 8 倍オーバーサンプリング処理と補間処理との組み合わせを示すものである。

【0047】ところで、各ハーモニクス（高調波）の位相を求める他の方法として、入力信号波形の 1 ピッチ周期分を切り出し、切り出された波形データにゼロ詰めを施して全体で  $2^1$  サンプル（ $N$  は整数、 $2^1$  は上記 1 ピッチ周期のサンプル数以上で例えば 256）とし、FFT 等の直交変換を施して、得られたデータの実部と虚部とを用いて逆正接 ( $\tan^{-1}$ ) を求める計算により位相を求め、この位相を補間処理して各高調波毎の位相を求めることが考えられる。この場合には、FFT 後の位相は上記ピッチ周波数とは無関係に表れるため、補間処理によって各ハーモニクス（高調波）毎の位相を計算することが必要とされる。

【0048】このような次元変換を施さずに FFT を施して後で補間処理して各ハーモニクスの位相を求める場合と、本実施の形態のように 1 ピッチ周期分の 20 ~ 147 のサンプルを  $2^1$  のサンプルに次元変換した後に F

$$pch / fs = 2^1 / fs' = 128 / fs' \quad (7)$$

であるから、ピッチ周期  $T$  に対応するピッチ周波数  $f$ 、

$$f = fs / pch = fs' / 2^1 = fs' / 128 \quad (8)$$

となる。この結果は、次元変換された図 1 1 の (C) の波形データを FFT 処理すれば、得られたスペクトルの間隔は、図 1 1 の (D) に示すように、 $f$  となることを意味している。従って、位相の場合も、先に次元変換して FFT することにより、直接的に各ハーモニクスの位相が求められることになる。

【0051】このように、FFT 前に次元変換を施しておくことにより、FFT 後に線形補間処理等を行うことなく直接的に各ハーモニクスの位相を求めることができるため、位相の補間処理の手間が省け、また補間による誤差も低減できる。

【0052】次に、上述のようにして求められた位相情報を用いてサイン波合成を行う場合の具体例について図 1 2 を参照しながら説明する。ここでは、時刻  $n_1$  から  $n_2$  までのフレーム間隔  $L = n_2 - n_1$  の時間波形をサイン波合成 (Sinusoidal 合成) により再生する場合について説明する。

$$A_m(n) = \frac{n_2 - n}{L} A_{1m} + \frac{n - n_1}{L} A_{2m}$$

【0056】時刻  $n_1, n_2$  の間での  $m$  番目のハーモニクス成分の周波数変化を、次の (10) 式で示すように、（線形変化分）+（固定変動分）であると仮定する。

FFT 処理して直接各ハーモニクスの位相を求める場合とを比較しながら説明する。図 1 1 は、この比較の説明に供する図であり、この図 1 1 では、いずれの場合も 128 サンプル FFT を行う例を示している。また、FFT 結果はスペクトルで表している。

【0049】図 1 1 の (A) は、入力信号から  $pch$  サンプル分を切り出したものを示しており、残りをゼロ詰めて 128 サンプル（一般には  $2^1$  サンプル）とした後に FFT 処理することにより、図 1 1 の (B) に示すようなスペクトルが得られる。このスペクトルの間隔は、サンプリング周波数を  $fs$  とするとき  $fs / 128$  となる。これに対して、 $pch$  サンプルを 128 サンプル（一般には  $2^1$  サンプル）に次元変換した図 1 1 の (C) の波形データは、次元変換によってサンプリング周波数が  $fs'$  に変化しており、これを FFT 処理するとスペクトル間隔は  $fs' / 128$  となる。この次元変換後のサンプリング周波数  $fs'$  について考察する。

【0050】先ず、ピッチラグが  $pch$  (サンプル) であるので、次元変換前のピッチ周期  $T$  は、

$$T = pch / fs \quad (5)$$

となり、このピッチ周期  $T$  と上記次元変換後のサンプリング周波数  $fs'$  との関係は、

$$T = 2^1 / fs' = 128 / fs' \quad (6)$$

となる。よって、

$$pch / fs = 2^1 / fs' = 128 / fs' \quad (7)$$

(Hz) は、

$$f = fs / pch = fs' / 2^1 = fs' / 128 \quad (8)$$

【0053】時刻  $n_1$  のピッチラグが  $pch_1$  (サンプル)、時刻  $n_2$  のピッチラグが  $pch_2$  (サンプル) であるとき、時刻  $n_1, n_2$  のピッチ周波数  $\omega_1, \omega_2$  (rad/サン

$$\omega_1 = 2\pi / pch_1$$

$$\omega_2 = 2\pi / pch_2$$

である。また、各ハーモニクス成分の振幅データを、時刻  $n_1$  では、 $A_{11}, A_{12}, A_{13}, \dots$ 、時刻  $n_2$  では、 $A_{21}, A_{22}, A_{23}, \dots$  とし、各ハーモニクス成分の位相データを時刻  $n_1$  では、 $\phi_{11}, \phi_{12}, \phi_{13}, \dots$ 、時刻  $n_2$  では、 $\phi_{21}, \phi_{22}, \phi_{23}, \dots$  とする。

【0054】ピッチが連続している場合には、時刻  $n'$  ( $n_1 \leq n \leq n_2$ ) における第  $m$  番目のハーモニクス成分の振幅は、時刻  $n_1, n_2$  における振幅データの線形補間によって、次の (9) 式により得られる。

【0055】

【数 3】

$$(n_1 \leq n \leq n_2) \quad (9)$$

【0057】

【数 4】



$$\omega_m(n) = m\omega_1 \frac{n_2 - n}{L} + m\omega_2 \frac{n - n_1}{L} + \Delta\omega_m \quad (n_1 \leq n \leq n_2) \quad (10)$$

【0058】このとき、第  $m$  番目のハーモニクス成分の時刻  $n$  における位相  $\theta_m(n)$  (rad) は、次の (12) 式で表されるから、これを計算して (13) 式が得られ

$$\theta_m(n) = \int_{n_1}^n \omega_m(\xi) d\xi + \phi_{1m} \quad (11)$$

$$= \int_{n_1}^n \left( m\omega_1 \frac{n_2 - \xi}{L} + m\omega_2 \frac{\xi - n_1}{L} + \Delta\omega_m \right) d\xi + \phi_{1m} \quad (12)$$

$$= m\omega_1(n - n_1) + m(\omega_2 - \omega_1) \frac{(n - n_1)^2}{2L} + \Delta\omega_m L + \phi_{1m} \quad (13)$$

【0060】よって、時刻  $n$  における  $m$  番目のハーモニクスの位相  $\phi_m(n)$  (rad) は、次の (15) 式で表される。従って各ハーモニクス成分の周波数変化の変動分  $\Delta\omega_m$  (rad/サン'ル) は、次の (16) 式に示すようになる。

【0061】

【数6】

$$\phi_{2m} = \theta_m(n_2) \quad (14)$$

$$= \frac{m(\omega_1 + \omega_2)L}{2} + \Delta\omega_m L + \phi_{1m} \quad (15)$$

$$W_m(n) = A_m(n) \cos(\theta_m(n))$$

となる。このようにして得られた全てのハーモニクスに関する時間波形の総和をとったものが、次の (18) 式、(19) 式に示すように、合成波形  $V(n)$  となる。

$$V(n) = \sum_m W_m(n) \quad (18)$$

$$= \sum_m A_m(n) \cos(\theta_m(n)) \quad (n_1 \leq n \leq n_2) \quad (19)$$

【0065】次に、ピッチ不連続の場合について説明する。ピッチ不連続の場合は、周波数変化の連続性は考慮せずに、時刻  $n_1$  より前向きにサイン波合成した次の (20) 式に示す波形  $V_1(n)$  と、時刻  $n_2$  より後ろ向き

$$V_1(n) = \sum_m A_{1m} \cos(m\omega_1(n - n_1) + \phi_{1m}) \quad (20)$$

【0067】

$$V_2(n) = \sum_m A_{2m} \cos(-m\omega_2(n_2 - n) + \phi_{2m}) \quad (21)$$

【0068】以上説明したような位相検出装置によれば、予め検出されたピッチ周波数を用いて、FFTと線形補間により、所望のハーモニクス成分の位相を高速に検出できる。これにより、音声信号のサイン波合成符号化、又は音声信号のLPC残差にサイン波合成符号化を用いる音声符号化において、波形再現性を実現できる。

【0069】なお、本発明は上記実施の形態のみに限定されるものではなく、例えば上記図1の構成については、各部をハードウェア的に記載しているが、いわゆるDSP（デジタル信号プロセッサ）等を用いてソフトウェアプログラムにより実現することも可能である。

る。

【0059】

【数5】

【0062】

【数7】

$$\Delta\omega_m = \frac{\phi_{1m} - \phi_{2m}}{L} - \frac{m(\omega_1 + \omega_2)}{2} \quad (16)$$

【0063】第  $m$  番目のハーモニクス成分について、時刻  $n_1, n_2$  における位相  $\phi_{1m}, \phi_{2m}$  が与えられているので、上記 (16) 式より、周波数変化の固定変動分  $\Delta\omega_m$  を求め、上記 (13) 式により各時刻  $n$  の位相  $\theta_m$  が求まれば、第  $m$  番目のハーモニクスによる時間波形  $W_m(n)$  は、

$$(n_1 \leq n \leq n_2) \quad (17)$$

【0064】

【数8】

にサイン波合成した次の (21) 式に示す波形  $V_2(n)$  とにそれぞれ窓をかけて重畳加算 (overlap add) する。

【0066】

【数9】

【数10】

【0070】

【発明の効果】以上の説明から明らかなように、本発明に係る位相検出装置及び方法によれば、音声信号に基づく入力信号波形を時間軸上で1ピッチ周期分だけ切り出し、切り出された1ピッチ周期分のサンプルに対して次元変換を施して  $2^L$  サンプルのデータにした後、 $2^L$  ポイントFFT等の直交変換を施し、直交変換されたデータの実部と虚部とを直接用いて上記入力信号の各高調波成分の位相情報を検出することにより、元の波形の位相情報を簡単に検出でき、波形再現性を高めることができる。

【0071】特に、予め検出されたピッチを用いて、次元変換とFFT（高速フーリエ変換）とを用いることにより、各ハーモニクス（高調波）成分の位相を高速に検出できる。これによって、サイン波合成符号化に適用した場合に、波形再現性を高めることができ、例えば合成音が不自然になることを未然に防止できる。

【図面の簡単な説明】

【図1】本発明に係る位相検出装置及び方法の実施の形態が適用される音声符号化装置の一例の概略構成を示すブロック図である。

【図2】本発明に係る実施の形態となる位相検出装置の概略構成を示すブロック図である。

【図3】本発明に係る実施の形態となる位相検出方法を説明するためのフローチャートである。

【図4】位相検出の対象となる入力信号の一例を示す波形図である。

【図5】1ピッチ分の波形及び次元変換後の波形の一例を示す波形図である。

【図6】検出された位相の一例を示す図である。

【図7】次元変換のためのオーバーサンプリングフィルタのフィルタ係数の一例を示す図である。

【図8】次元変換のためのオーバーサンプリングフィルタのフィルタ特性の一例を示す図である。

【図9】次元変換処理に用いられる変数が蓄えられるメモリの一例を示す図である。

【図10】次元変換の処理手順の一例を説明するためのフローチャートである。

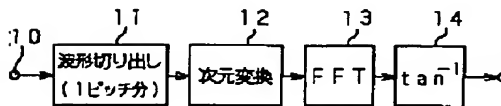
【図11】FFT処理前に次元変換を行うことの意味を説明するための図である。

【図12】位相情報が得られたときのサイン波合成の一例を説明するための図である。

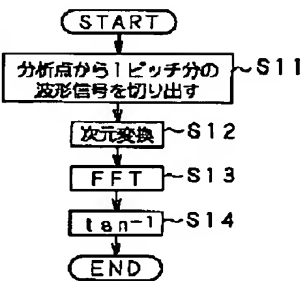
【符号の説明】

11 波形切り出し部、 12 次元変換部、 13 FFT処理部、 14  $\tan^{-1}$ 部、 110 第1の符号化部、 111 オーブンループピッチサーチ部、 112 直交変換部、 113 高精度ピッチサーチ部、 114 V/U/V判定部、 115 スペクトルエンベロープ評価部、 116 スペクトルエンベロープ量子化部、 120 第2の符号化部、 131 LPC逆フィルタ、 132 LPC分析部、 133 LSP量子化部、 141 位相検出部、 142 位相量子化部

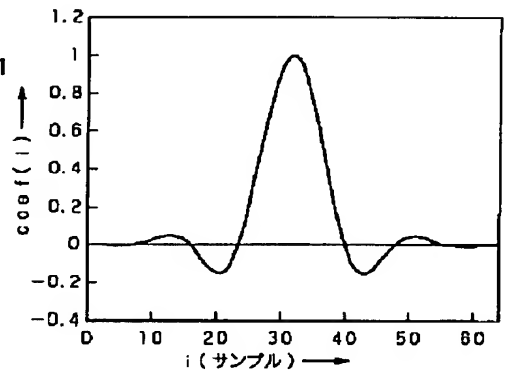
【図2】



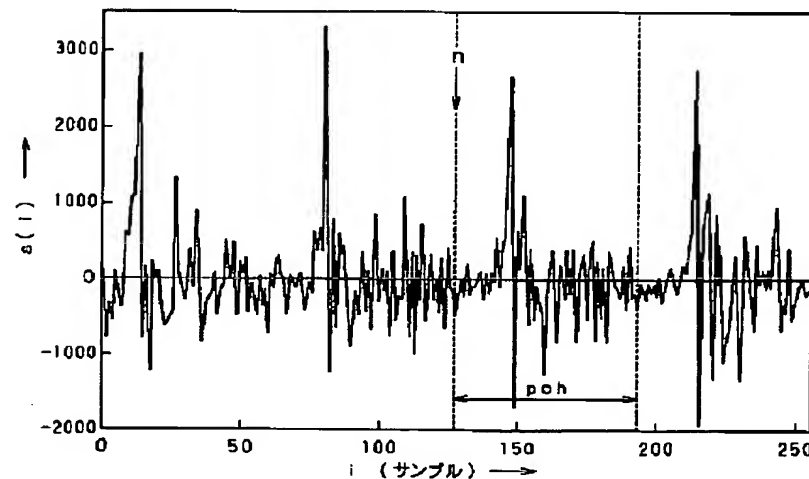
【図3】



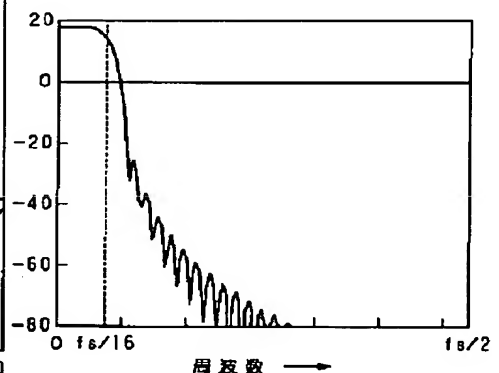
【図7】



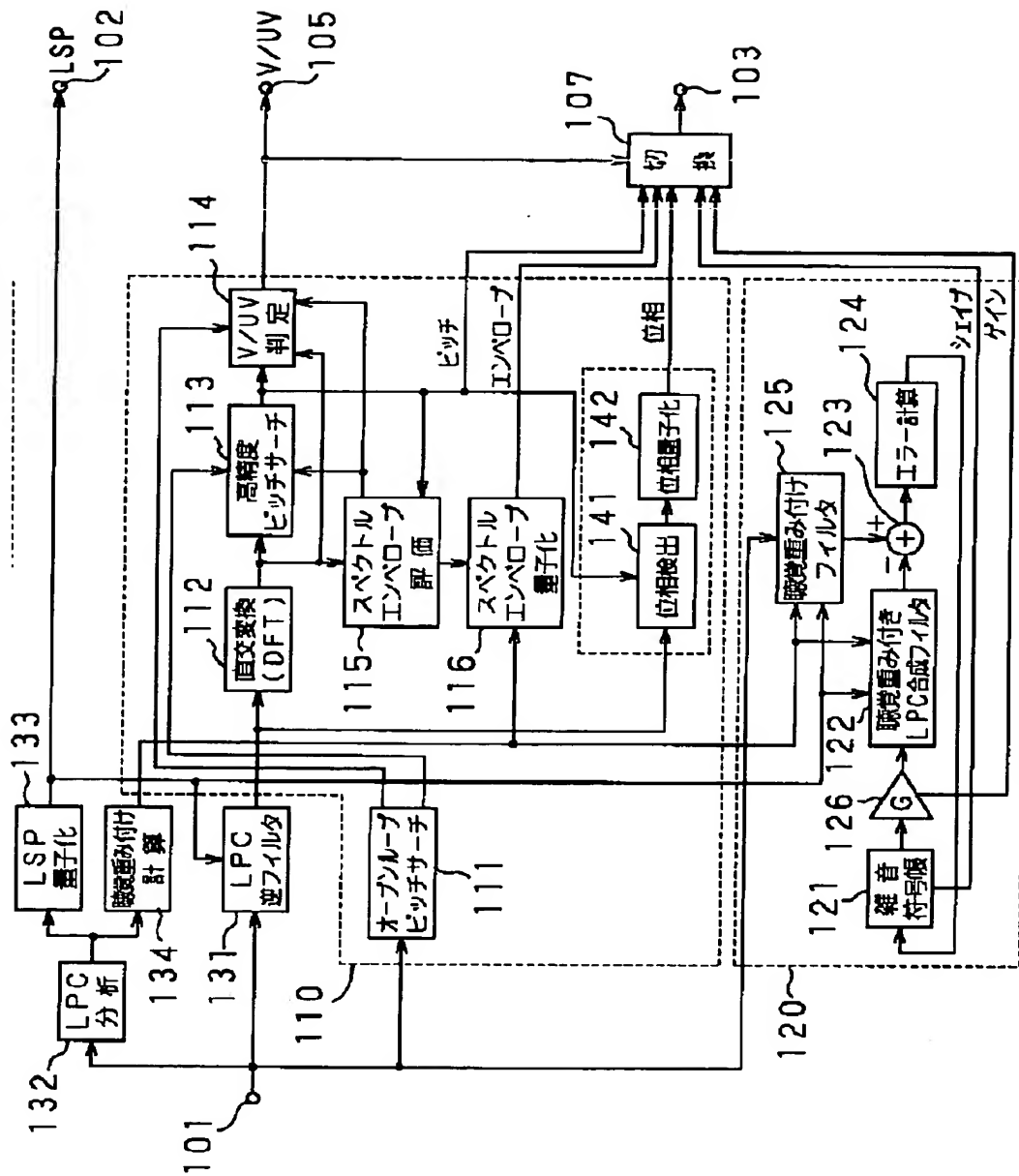
【図4】



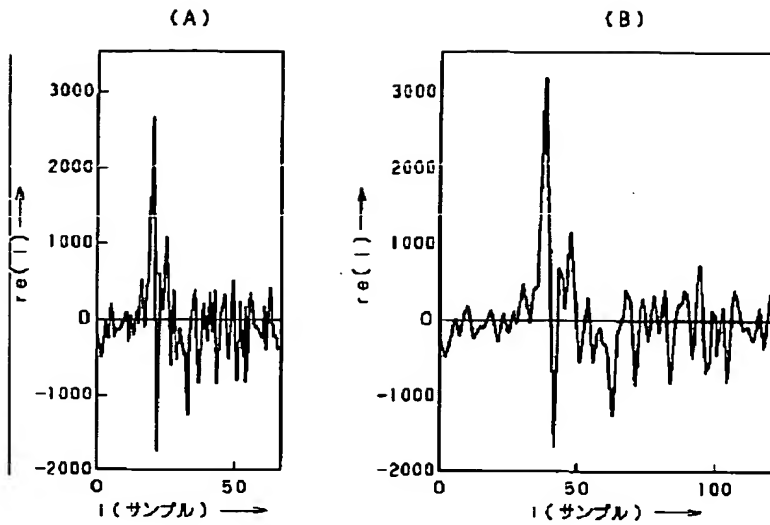
【図8】



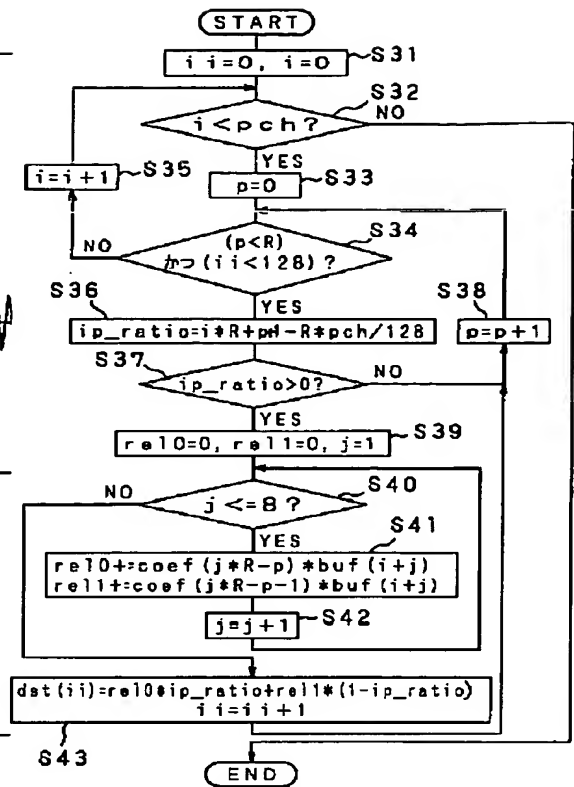
【 図 1 】



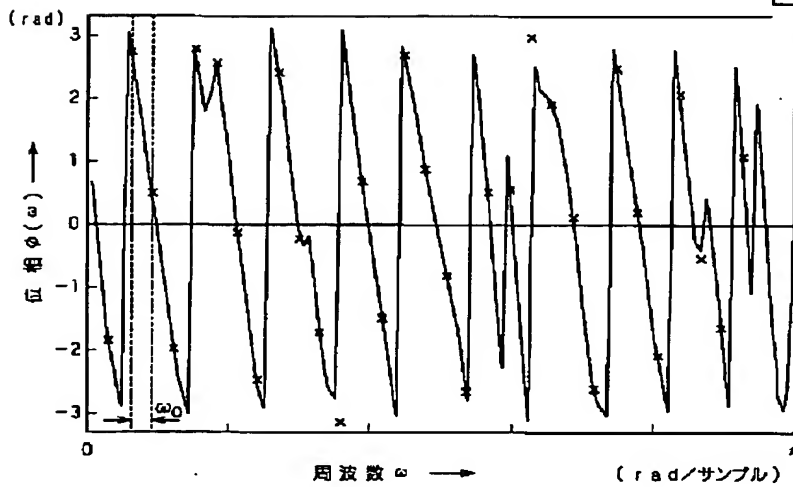
【図 5】



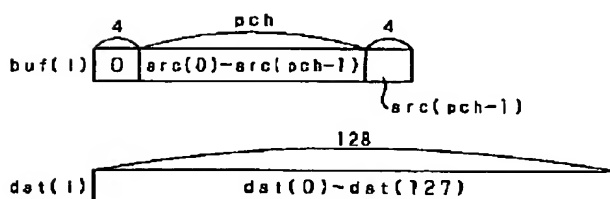
【図 10】



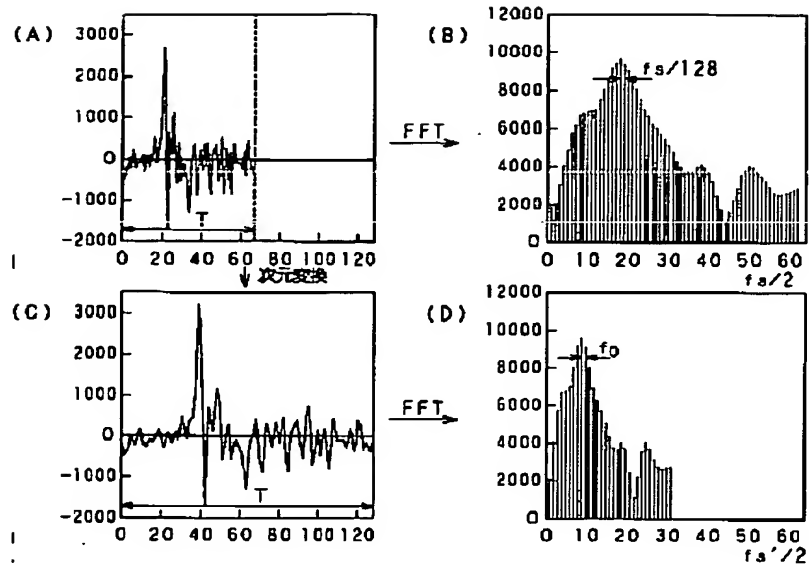
【図 6】



【図 9】



【 図 1 1 】



【 図 1 2 】

